METHOD FOR BUILDING SPACE-SPLITTING DECISION TREE

Abstract

A method is provided for data classification that achieves improved
interpretability and accuracy while preserving the efficiency and scalability of

5   univariate decision trees. To build a compact decision tree, the method
searches for clusters in subspaces to enable multivariate splitting based on
weighted distances to such a cluster. To classify an instance more accurately,
the method performs a nearest neighbor (NN) search among the potential
nearest leaf nodes of the instance. The similarity measure used in the NN

10   search is based on Euclidean distances defined in different subspaces for
different leaf nodes. Since instances are scored by their similarity to a certain
class, this approach provides an effective means for target selection that is not
supported well by conventional decision trees.